

# Bayesian Inference on Parametric Mixture Cure Model with Spatial Autocorrelation in Random Effects

Bomin Kim

July 1, 2017

## **Abstract**

In this paper, we implement Bayesian inference on parametric mixture cure model, one of the most popular models to estimate the cure rate of treatment and the survival rate of uncured patients at the same time, with spatial autocorrelation in random effects. This includes two parametric survival models, Exponential and Weibull.

# Review on Parametric Mixture Cure Model

## *Likelihood function*

We consider a model for the duration  $t$  which splits the sample into two groups, one of which will eventually experience the event of interest (i.e., “fail”) and the other which will not. We define the latent variable  $Y$  such that  $Y_i = 1$  for those who will eventually fail and  $Y_i = 0$  for those who will not; define  $Pr(Y_i = 1) = \delta_i$  such that the probability  $\delta_i$  is modeled as a logit by including explanatory variables:

$$\delta_i = \frac{\exp(Z_i\gamma)}{1 + \exp(Z_i\gamma)}. \quad (1)$$

The likelihood function derived by Box-Steffensmeier and Zorn (1999) is:

$$L = \prod_{i=1}^N [\delta_i g(t_i|X_i, \beta)]^{C_i} [1 - \delta_i + \delta_i G(t_i|X_i, \beta)]^{1-C_i}, \quad (2)$$

where  $g(t_i|X_i, \beta) = f(t_i * Y_i = 1|X_i, \beta)$  and  $G(t_i|X_i, \beta) = S(t_i * Y_i = 1|X_i, \beta)$ . Here, we follow the common definition in survival model literature such that  $f(t_i|X_i, \beta)$  is the density function and  $S(t_i|X_i, \beta) = Pr(T_i > t_i|X_i, \beta)$ , with  $\beta$  and  $\gamma$  are the parameter vectors to be estimated.

Then the corresponding log-likelihood function is

$$\ln L = \sum_{i=1}^N C_i [\ln \delta_i + \ln g(t_i|X_i, \beta)] + (1 - C_i) \ln [1 - \delta_i + \delta_i G(t_i|X_i, \beta)]. \quad (3)$$

We will expand this mixture cure model to be the spatial-cure model, by incorporating spatial autocorrelation as a random effect.

## Derivation of Parametric Spatial Mixture Cure Model

Including a random effect or “frailty” term, we define the proportional hazards  $h(t_i|X_i, \beta)$  as:

$$\begin{aligned} h(t_i|X_i, \beta) &= h_0(t_i)\omega_i \exp(X_i\beta) \\ &= h_0(t_i) \exp(X_i\beta + W_i), \end{aligned} \tag{4}$$

where  $h_0(t_i)$  is the baseline hazard and  $W_i \equiv \ln \omega_i$  is the individual frailty term, designed to capture differences among the individuals.

For cure model, we can also add the random effect term to  $\Pr(Y_i = 1)$  such that

$$\delta_i = \frac{\exp(Z_i\gamma + W_i)}{1 + \exp(Z_i\gamma + W_i)}. \tag{5}$$

Should we use shared random effect  $W_i$ ? or define different random effect  $V_i$ ? (as we did for  $\gamma$ )

Then, we can define the likelihood functions for Exponential and Weibull models:

### 1. Exponential

$$\begin{aligned} f(t_i|X_i, \beta, W_i) &= \exp(X_i\beta + W_i)\exp(-\exp(X_i\beta + W_i)t_i), \\ S(t_i|X_i, \beta, W_i) &= \exp(-\exp(X_i\beta + W_i)t_i), \end{aligned} \tag{6}$$

which establish the likelihood and log-likelihood function respectively as

$$\begin{aligned} L(\beta, \gamma, \mathbf{W}) &= \prod_{i=1}^N [\delta_i \exp(X_i\beta + W_i) \exp(-\exp(X_i\beta + W_i)t_i)]^{C_i} \\ &\quad \times [1 - \delta_i + \delta_i \exp(-\exp(X_i\beta + W_i)t_i)]^{1-C_i}, \end{aligned} \tag{7}$$

$$\begin{aligned} \ln L(\beta, \gamma, \mathbf{W}) &= \sum_{i=1}^N C_i [\ln \delta_i + (X_i\beta + W_i) - \exp(X_i\beta + W_i)t_i] \\ &\quad + (1 - C_i) \ln [1 - \delta_i + \delta_i \exp(-\exp(X_i\beta + W_i)t_i)]. \end{aligned} \tag{8}$$

## 2. Weibull

$$\begin{aligned}
f(t_i|\rho, X_i, \beta, W_i) &= \rho(\exp(X_i\beta + W_i))(\exp(X_i\beta + W_i)t_i)^{\rho-1}\exp(-(\exp(X_i\beta + W_i)t_i)^\rho) \\
S(t_i|\rho, X_i, \beta, W_i) &= \exp(-(\exp(X_i\beta + W_i)t_i)^\rho),
\end{aligned} \tag{9}$$

which establish the likelihood and log-likelihood function respectively as

$$\begin{aligned}
L(\rho, \beta, \gamma, \mathbf{W}) &= \prod_{i=1}^N [\delta_i \rho(\exp(X_i\beta + W_i))(\exp(X_i\beta + W_i)t_i)^{\rho-1}\exp(-(\exp(X_i\beta + W_i)t_i)^\rho)]^{C_i} \\
&\quad \times [1 - \delta_i + \delta_i \exp(-(\exp(X_i\beta + W_i)t_i)^\rho)]^{1-C_i},
\end{aligned} \tag{10}$$

$$\begin{aligned}
\ln L(\rho, \beta, \gamma, \mathbf{W}) &= \sum_{i=1}^N C_i [\ln \delta_i + \ln \rho + (X_i\beta + W_i) + (\rho - 1)(X_i\beta + W_i + \ln t_i) \\
&\quad - \rho \exp(X_i\beta + W_i)t_i] + (1 - C_i) \ln [1 - \delta_i + \delta_i \exp(-(\exp(X_i\beta + W_i)t_i)^\rho)].
\end{aligned} \tag{11}$$

For Bayesian inference, we use simplest specifications for the priors, for example,:

$$\begin{aligned}
\beta &\sim \text{Multivariate Normal}_{p_1}(\boldsymbol{\mu}_\beta, \Sigma_\beta), \\
\gamma &\sim \text{Multivariate Normal}_{p_2}(\boldsymbol{\mu}_\gamma, \Sigma_\gamma), \\
\rho &\sim \text{Gamma}(a_\rho, b_\rho) \text{ for Weibull.}
\end{aligned} \tag{12}$$

For the hyperparameters, we can follow the common approach and use

$$\begin{aligned}
\boldsymbol{\mu}_\beta &= \mathbf{0}, \Sigma_\beta \sim \text{Inverse-Wishart}(p_1 \mathbf{I}_{p_1}, p_1), \\
\boldsymbol{\mu}_\gamma &= \mathbf{0}, \Sigma_\gamma \sim \text{Inverse-Wishart}(p_2 \mathbf{I}_{p_2}, p_2), \\
(a_\rho, b_\rho) &= (0.001, 1/0.001) = (0.001, 1000),
\end{aligned} \tag{13}$$

where we use hierarchical Bayesian modeling to estimate  $\Sigma_\beta$  and  $\Sigma_\gamma$  using Inverse-Wishart distribution. Note that if this step seems to be unnecessary, we can instead simply fix those such as  $\Sigma_\beta = \Sigma_\gamma = 10^4 \times \mathbf{I}$  (very slow mixing).

## Spatial Modeling with Conditionally Autoregressive (CAR) model

Now, we need to define the priors for the random effect  $\mathbf{W} = \{W_1, \dots, W_N\}$  to assign spatial correlation across the individuals. We follow the canonical approach and apply conditionally autoregressive (CAR) model, one of the most widely used model in spatial statistics.

To employ spatial process modeling for the geographic locations, we define our random effect to be naturally associated with areal units. As such we work exclusively with CAR models for these effects, i.e., we assume that

$$\mathbf{W}|\lambda \sim \text{CAR}(\lambda), \quad (14)$$

where  $\lambda$  refers to the precision of the random effects distribution of the random effects distribution. By allowing this parameter, we can avoid the exchangeable prior and reflect a conditionally autoregressive prior that incorporates neighbor definitions via the spatial weights matrix.

### 1. Individual frailty model

Similarly as non-spatial frailty (or random effect) models, this model assumes that the individual unit  $i$  has spatial-related random effect,  $W_i$ . The conditional distribution of the spatial random effects that results from individual frailty model CAR prior is then:

$$W_i|W_{i' \neq i} \sim N(\overline{W}_i, 1/(\lambda m_i)), \quad (15)$$

where  $\overline{W}_i$  is the average of the  $W_{i' \neq i}$  neighboring  $W_i$  (i.e.  $\overline{W}_i = m_i^{-1} \sum_j \text{adj}_i W_j$ ), and  $m_i$  is the number of adjacencies. By incorporating the spatial locations of units, the CAR prior thus produces a conditional distribution for the random effects that is normally distributed with a conditional mean equal to the average of the random effects for neighbors of  $i$ , and a conditional variance that is inversely proportional to the number of units neighboring  $i$  (Thomas et al. 2004).

## 2. Shared frailty model

This model assumes that the individual unit  $i$  is now nested in a higher-level cluster or stratum  $j$ , and the random effect refers to this higher-level stratum,  $W_j$ . The resulting conditional distribution for the strata-level spatial random effects is then:

$$W_j | W_{j' \neq j} \sim N(\overline{W}_j, 1/(\lambda m_j)), \quad (16)$$

where  $\overline{W}_j$  is the average of the  $W_{j' \neq j}$  neighboring  $W_j$  (i.e.  $\overline{W}_j = m_j^{-1} \sum_k \text{adj}_j W_k$ ), and  $m_j$  is the number of adjacencies. Analogous to the individual frailty case, the CAR prior thus produces a conditional distribution for the spatial shared frailties that is normally distributed with a conditional mean equal to the average of the random effects for strata neighboring stratum  $j$ , and a conditional variance that is inversely proportional to the number of strata neighboring  $j$  (Thomas et al. 2004).

The individual and shared spatial frailty models also require that a hyperprior,  $p(\lambda)$ , be assigned to  $\lambda$ . Generally, Gamma hyperprior is chosen (Banerjee, Carlin, and Gelfand 2004) as

$$\lambda \sim \text{Gamma}(a_\lambda, b_\lambda), \quad (17)$$

where we can specify vague prior  $(a_\lambda, b_\lambda) = (0.001, 1/0.001) = (0.001, 1000)$  as we did for  $\rho$ . Moreover, to identify an intercept, we may impose the constraint that the frailties sum to zero, i.e.  $\sum_i W_i = 0$ .

## Inference

Now, the joint posterior distribution for the Bayesian parametric (Weibull, but reduces to Exponential with  $\rho = 1$ ) spatial mixture cure model is:

$$\begin{aligned} & \pi(\beta, \gamma, \rho, \mathbf{W}, \lambda, \Sigma_\beta, \Sigma_\gamma | \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) \\ & \propto L(\beta, \gamma, \rho, \mathbf{W} | \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) \pi(\mathbf{W} | \lambda) \pi(\beta | \boldsymbol{\mu}_\beta, \Sigma_\beta) \pi(\gamma | \boldsymbol{\mu}_\gamma, \Sigma_\gamma) \pi(\rho) \pi(\lambda) \pi(\Sigma_\beta) \pi(\Sigma_\gamma), \end{aligned} \quad (18)$$

where  $L(\beta, \gamma, \rho, \mathbf{W} | \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t})$  is Equation (10),  $\pi(\mathbf{W} | \lambda)$  is either Equation (15) or Equation (16) based on the choice of individual or shared frailty model,  $\pi(\beta | \boldsymbol{\mu}_\beta, \Sigma_\beta)$  and  $\pi(\gamma | \boldsymbol{\mu}_\gamma, \Sigma_\gamma)$  comes from Equation (12),  $\pi(\rho)$  and  $\pi(\lambda)$  are Gamma hyperpriors in Equation (12) and Equation (17), and  $\pi(\Sigma_\beta)$  and  $\pi(\Sigma_\gamma)$  are from Equation (13).

For Bayesian inference, we will implement Markov Chain Monte Carlo (MCMC) algorithm with the following update scheme:

**Step 0.** Choose an arbitrary starting point  $\beta_0, \gamma_0, \rho_0$  if Weibull,  $\lambda_0$ , and corresponding  $\mathbf{W}_0 = \{W_1, \dots, W_N\}$  from the prior given  $\lambda_0$ ; then set  $i = 0$ .

**Step 1.** Update  $\Sigma_\beta \sim \pi(\Sigma_\beta | \beta)$ ,  $\Sigma_\gamma \sim \pi(\Sigma_\gamma | \gamma)$ , and  $\lambda \sim \pi(\lambda | \mathbf{W})$  using Gibbs sampler.

**Step 2.** Update  $\beta \sim \pi(\beta | \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{W}, \gamma, \rho, \boldsymbol{\mu}_\beta, \Sigma_\beta)$ ,  $\gamma \sim \pi(\gamma | \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{W}, \beta, \rho, \boldsymbol{\mu}_\gamma, \Sigma_\gamma)$ , and  $\mathbf{W} \sim \pi(\mathbf{W} | \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \beta, \gamma, \rho, \lambda)$  using slice sampling.

**Step 2'.** If Weibull, update  $\rho \sim \pi(\rho | \mathbf{C}, \boldsymbol{\alpha}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \mathbf{W}, \beta, \gamma, a_\lambda, b_\lambda)$  using slice sampling.

**Step 3.** Set  $i = i + 1$ , and go to Step 1.

**Step 4.** After  $N$  iterations, summarize the parameter estimates using all sampled values (e.g. confidence intervals for posterior estimates)

## Appendix

While we already derived  $\pi(\Sigma_\beta|\boldsymbol{\beta}_i)$  and  $\pi(\Sigma_\gamma|\boldsymbol{\gamma}_i)$  (See BayesianZombieSurvivalModel draft), the closed form of full conditional distributions of  $\lambda \sim \pi(\lambda|\mathbf{W})$  in Step 1 are derived as below:

$$\begin{aligned}
\pi(\lambda|\mathbf{W}) &\propto \pi(\mathbf{W}|\lambda) \times \pi(\lambda) \\
&\propto \left( \prod_{i=1}^N \pi(W_i|\lambda) \right) \times \pi(\lambda) \\
&\propto \left( \prod_{i=1}^N N(\overline{W}_i, 1/(\lambda m_i)) \right) \times \text{Gamma}(a_\lambda, b_\lambda) \\
&\propto \lambda^{\frac{N}{2}} \exp\left\{-\frac{\lambda m_i}{2} \sum_{i=1}^N (W_i - \overline{W}_i)^2\right\} \times \lambda^{a_\lambda-1} \exp\{-b_\lambda \lambda\} \\
&\propto \lambda^{\frac{N}{2}+a_\lambda-1} \exp\left\{-\left(\frac{m_i}{2} \sum_{i=1}^N (W_i - \overline{W}_i)^2 + b_\lambda\right) \lambda\right\} \\
&\sim \text{Gamma}\left(\frac{N}{2} + a_\lambda, \quad \frac{m_i}{2} \sum_{i=1}^N (W_i - \overline{W}_i)^2 + b_\lambda\right)
\end{aligned}$$